CLAIMS

What is claimed is:

1. A computerized method of generating a data mining model, the method comprising:

obtaining objectives for the data mining model;

automatically selecting a set of algorithms based on the objectives;

obtaining sample data;

creating a plurality of datasets from the sample data;

optimizing the set of algorithms using the plurality of datasets; and

generating the data mining model based on the optimized set of algorithms.

2. The method of claim 1, wherein the creating step includes:

shuffling the sample data;

placing the shuffled sample data into a plurality of partitions; and

including each partition in one of the plurality of datasets.

3. The method of claim 2, wherein the plurality of datasets includes a training dataset, a validation dataset, and a testing dataset.

4. The method of claim 3, wherein the creating step further includes repeating the including step until each partition is included in at least one training dataset.

5. The method of claim 1, wherein the selecting step includes obtaining a rule that comprises a best practice for an objective.

6. The method of claim 5, wherein the best practice is based on at least one of: research, data characteristics, and user feedback.

7. The method of claim 1, wherein the selecting step includes analyzing an attribute of the sample data, and wherein the set of algorithms is further selected based on the attribute.

8. The method of claim 1, wherein the optimizing step includes:

applying the set of algorithms to the plurality of datasets; and

analyzing a set of results for the applying step.

9. The method of claim 8, wherein the optimizing step further includes:

adjusting at least one algorithm based on the set of results; and

applying the adjusted set of algorithms to the plurality of datasets.

10. The method of claim 1, wherein the generating step includes generating a set of standard query language statements based on the optimized set of algorithms.

11. The method of claim 1, further comprising storing the data mining model as a character large object (CLOB) in a database.

12. A computerized method of generating a data mining model, the method comprising:

obtaining a set of algorithms and a plurality of datasets;

applying the set of algorithms to the plurality of datasets;

analyzing a set of results for the applying step;

adjusting at least one algorithm based on the set of results;

applying the adjusted set of algorithms to the plurality of datasets; and

generating the data mining model based on the adjusted set of algorithms.

13. The method of claim 12, wherein the obtaining step includes:

obtaining sample data; and

automatically generating the plurality of datasets from the sample data.

14. The method of claim 12, wherein the obtaining step includes:

obtaining objectives for the data mining model; and

automatically selecting the set of algorithms based on the objectives.

15. A system for generating a data mining model, the system comprising:

a dataset system for obtaining a plurality of datasets;

a rules system for obtaining a plurality of algorithms;

an optimization system for optimizing the set of algorithms using the plurality of datasets; and

a model system for generating the data mining model based on the optimized set of algorithms.

16. The system of claim 15, further comprising a storage system for storing the data mining model in a database.

17. The system of claim 15, wherein the dataset system automatically generates the plurality of datasets from sample data.

18. The system of claim 15, wherein the rules system automatically selects the set of algorithms based on objectives for the data mining model.

19. A program product stored on a recordable medium for generating a data mining model, which when executed comprises:

program code for generating a plurality of datasets from sample data;

program code for selecting a set of algorithms based on objectives for the data mining model;

program code for optimizing the set of algorithms using the plurality of datasets; and

program code for generating the data mining model based on the optimized set of algorithms.

20. The program product of claim 19, further comprising program code for storing the data mining model as a character large object (CLOB) in a database.

21. The program product of claim 19, wherein the program code for generating the data mining model includes program code for generating a set of standard query language statements based on the optimized set of algorithms.

22. The program product of claim 19, wherein the program code for generating the plurality of datasets includes:

program code for shuffling the sample data;

program code for placing the shuffled sample data into a plurality of partitions; and

program code for including each partition in one of the plurality of datasets.